DUPLICATE

Forecasting Development

Forecasting Research Technical Report No. 381

Forecasting Techniques Review Project

May 2002

P.F. Lavington

© Crown Copyright 2002

Met Office , Forecasting Development , Aviation Technology Centre, Room R321 , London Road , Bracknell , Berkshire ,RG12 2SZ, United Kingdom

00/0486

Forecasting Research
Technical Report No. 381

Forecasting Techniques Review Project

By

P.F. Lavington

May 2002

The Met Office
Forecasting Development
Room R307
London Road
Bracknell
RG12 2SZ
United Kingdom

© Crown Copyright 2002

Permission to quote from this paper should be obtained from the above Met Office division

Please notify us if you change your address or no longer wish to receive these publications.

Tel: 44 (0)1344 856245 Fax: 44 (0)1344 854026 email: jsarmstrong@meto.gov.uk

Forecasting Techniques Review Project

Abstract: The project produced clear evidence that for forecasting maximum and minimum temperatures the Single Site Forecasting Model (SSFM) and Model Output Statistics (MOS) are superior to traditional forecasting methods. The application of local experience from the forecaster, however, does produce a small improvement in accuracy over the traditional methods. Data analysed for other weather elements was generally too small for any definite conclusions to be made and in some cases verification proved very difficult. The performance of traditional methods and model products to predict fog, snow, low cloud, thunder and gusts were shown to be generally of limited help to the forecaster.

1. Introduction

A number of factors combined to make this an appropriate time for the instigation of a review of forecasting techniques currently being used. The primary factors were:

- ISO9000. In ISO9000 terms, forecasts are to be produced by 'appropriately trained personnel'. The Review is intended to investigate and check that forecasters are actually using the 'best' methods to produce their forecasts.
- Passage of time. As model products improve over time so reliance on their output increases. The Review investigates if the use of some of the traditional forecasting methods is still applicable.
- New scientific ideas. As new and sometimes different observational data become available, old forecasting techniques may no longer have the required input data available to them. The Review looks at the impact of these changes.

The outcome is to provide the basis of future developments of Met Office College forecasting courses and also produce useful data for modellers to check the performance of their products.

2. Project outline

2.1 Method

To implement the Project there were four steps to follow.

a) Establish which forecasting methods were being used and how popular they were.
b) Decide which methods were to be used in trials
c) Set up the trials and analyse the results over a period of at least six months.
d) Interpret the results and make recommendations. In addition, throughout the project, ideas were solicited from forecasters on any new techniques by e-mail and by using dedicated newsgroups for the Project.

2.2 Survey

In order to ascertain which forecasting methods were being used a survey was sent out to some 40 stations and replies were received from147 forecasters. The methods were those that appear in the Forecaster's Reference Book (FRB) and the survey returns provided statistics of their use, ranging through 'Frequently', 'Sometimes', to 'Never'.

The full results of the survey are given in *Appendix 1*

The main conclusions were that, out of the numerous forecasting methods in FRB, McKenzie's minimum temperature and Saunders fog point stand out as the only two that most forecasters regularly use. However, it has to be accepted that many of the methods for less common events, snow etc are perceived by forecasters to be used as frequently (or not) as the event itself! The survey results raised many questions but not necessarily clear cut answers. One had to start with the view that, if a method was used frequently by a large majority of forecasters, then that method must have stood the test of time and was considered effective. The 'Never' category was more difficult to interpret. Was the method not being used because the forecaster had no faith in its abilities or, alternatively, was it cumbersome to implement and forecasters had neither the time nor the inclination to use it? The view was taken that if 'Blogg's' method is in the FRB then 'Mr. Bloggs' must have produced convincing evidence of its effectiveness for it to have been published. Thus, if a method was cumbersome to use then that was the prime reason for it's lack of popularity rather than necessarily any idea that it was less than effective. When picking methods to test, the degree of difficulty of use was not considered as a factor to disregard that method.

2.3 The Trials

Many of the methods described in the FRB are variations on a theme – additions and embellishments to, basically, the same original method. Some used T-Phi constructions and others used regression equations. In deciding which methods to use in trials I looked at them taking into account their individuality along with popularity. If method A for, say Fog, used a representative T-Phi I took the most popular one, with or without additions, followed by the next most popular using a different approach, say a regression equation. This is especially important these days, as a constant complaint from forecasters is the lack of a representative ascent. By their nature, regression equations are often cumbersome and forecasters may well not use these methods for this very reason. They may, however, be as good as or even better than using simple T-Phi construction methods.

Decisions on which stations to take part in trials were based on trying to get as many forecasters involved as possible but there were constraints on using certain stations due to the unavailability of verifying data and other reasons. Type of station was also considered. For example, a civil station would have little or no customer demand for, say cloud base, whilst for a military station

it would be paramount. The elements were matched for station type and, as far as possible, geographic spread.

33 Stations in the UK took part with varying degrees of enthusiasm and the keenness of the staff involved was a factor in the amount of data received. However, what was received was generally enough for an analysis of some kind and to draw some valid statistical conclusions. 6 weather elements were tested with a selection of stations involved in one or, for some, two. The best selection was made to encompass various types and locations of stations.

A major problem was soon encountered regarding the number of events for certain weather elements. Obviously, max and min temperatures occurred every day but fog and snow events were much more difficult to capture and it remains questionable whether any useful statistics can be derived from the limited amount of data obtained. It is also recognised that there is a degree of bias in the data due to the fact that only occasions when the event actually occurred were analysed and no account was taken when the method predicted an event which did not occur. This was necessitated by practical considerations of the trials.

In designing the form of the trials I was acutely aware that co-operation depended on the ease of data gathering. Busy forecasters could not be expected to spend long periods of their time attending to a non-operational matter. Thus, I aimed to ensure that the entering of data for the trials should take no more than five minutes a day. Methods that used regression equations could fairly easily have them pre-programed into an Excel spreadsheet leaving the forecaster to enter a minimal amount of data. Data entered was, for example in the case of minimum temperature, the forecast mins from the two traditional methods, any 'local' method and the model product min temperatures from MOS and the SSFM. (MOS, though strictly not 'pure' model is treated as so throughout this report). In an attempt to capture the airmass involved, forecasters were asked to enter a Lamb type. (See *appendix 2* for detail). Obviously, the next day, the actual recorded min temperature was entered. Whilst a temperature trial was fairly straight forward, testing elements like fog and snow was more tricky. These required entering yes/no questions and certainly allowed a fair latitude of interpretation by the forecaster. i.e. was the snow symbol on the PGM exactly over his/her station? For the SSFM, by definition – site specific – there was a clear-cut yes/no answer. It was also stressed that nobody was checking up on the ability of the individual forecaster. All that was required was that the methods were tested and that we were not interested in what the forecaster actually issued on the day.

2.4 The Methods

2 or 3 traditional methods were used in each of the trials along with, what was referred to, as a local method. This was not specifically defined. It had been hoped that it would be a locally derived method applicable to the

station concerned and had been developed by local forecasters over time but never published in the FRB. However, this was not the outcome. It appeared that what this 'method' in truth is, is the undefinable forecaster experience factor. It probably varied in it's execution but ran along the lines of the forecaster using a mix of the traditional method, the model product, or his/her own local knowledge to modify the prediction. In para 3.2 of this paper it will be shown if this 'method' has it's attractions. Along with testing the performance of the traditional methods the forecast produced by model products was also entered. This was incorporated in the Project in order to provide an answer to the question, 'are forecasters using the best method to produce their forecasts'. If traditional method X consistently turns out to be inferior to a model product it is not unreasonable to question why the forecaster is not using the better model product. This will be addressed later in this report.

2.5 Interpretation of data

The trials started during November and December 2000. At the same time I made some trips to the outstations concerned to check all was being understood and no problems were encountered. As the data were returned the method of analysis had to be considered. To answer the question 'Which is the best method to use to forecast xxx?', is not entirely straightforward. The question raises many others. The simplistic answer is the one that, in the long run, comes closest to whatever one is forecasting. In terms of temperature, if method X has a Root Mean Square Error (RMS) of 1.8 °C from the actual temperature and method Y scores 2.1 °C then apparently method X is 'better' than method Y. However, one has to look at the spread of results. Method X may occasionally fail spectacularly, which could cause the forecaster embarrassment whilst Method Y may plod along and, although in the long run, not appear as good as Method X, it is never miles out.
A minimum of 20 observations/events are required to have any statistical significance, and with 100 events the 95% confidence limits were well in sight. It was decided to present the data in various statistical forms – these being:

a) Tables showing Root Mean Squares (RMS), Standard Deviations (SD) and Mean Errors (ME) of the difference between the forecast and the outturn.
b) Frequency distributions of the difference between the forecast and the outturn
c) Scatter Diagrams and Correlations

In all cases the error was defined as the forecast – actual. For example – forecast of 10 °C , actual of 8 °C – error is +2 °C .

The decision of whether to combine data from different stations or treat them as individuals was debatable. One could argue, a method is a method and it matters not where it is applied. However, it is possible that method A works better at station B and vice versa. Thus, data was presented in both forms. In

order to look for any signal other considerations were applied, with various degrees of success e.g. Airmass.

In the following sections the results are presented with the merits or otherwise of the traditional methods being mainly discussed. If appropriate the 'local' method and the model products will also be considered. There now follows a rundown of the results for all of the weather elements. They will be described in turn with the full explanation of methods of interpretation given mainly for the first element, maximum temperature. In section 4 there will be a general discussion of all the results combined with conclusions and any recommendations. Section 5 will contain comments on implications for training with final thoughts and acknowledgements in section 6.

3. Trial results for various weather elements

3.1 Maximum Temperature

4 Stations took part – Leeds, Birmingham, Aldergrove and Lossiemouth. Whilst not perfect these represented two inland stations and two coastal and a total of 705 observations were obtained. Two traditional methods to forecast maximum temperatures were tested along with a local method and two model products, MOS and SSFM. The two traditional methods used were Johnson & Jefferson (J&J) and Callen & Prescott (C&P). Both are based on a representative ascent with the former method relating the amount of solar energy available for heating to an equivalent area on the T-Phi and the latter using a regression equation incorporating the 1000-850mb thickness layer. (*Full details can be found in the FRB at 2.3.1 and 2.3.3*).

Although the two methods are fundamentally different, the RMS errors are very similar at all of the stations involved. J&J appeared marginally better overall. Mean errors (forecast-actual) followed a similar pattern with both methods tending to show a positive bias i.e. forecast a maximum temperature higher than actual. Full results are shown in Table 1

RMS Error		Methods				
J&J	C&P	Local	MOS	SSFM	Station	Obs
1.98	1.76	1.23	1.67	1.49	**Belfast**	156
1.59	1.61	1.42	1.71	N/A	**Leeds**	194
2.57	2.81	1.72	1.83	2.01	**Lossiemouth**	67
1.48	1.64	1.30	1.93	1.47	**Birmingham**	288
1.91	**1.95**	**1.42**	**1.79**	**1.66**	Mean	705

Mean Error		Methods				
J&J	C&P	Local	MOS	SSFM	Station	Obs
0.91	0.75	0.04	-0.10	-0.39	**Belfast**	156
-0.36	-0.11	-0.39	-0.26	N/A	**Leeds**	194
0.63	1.00	-0.03	-0.16	-1.00	**Lossiemouth**	67
0.16	0.21	0.00	-0.06	-0.06	**Birmingham**	288
0.33	**0.46**	**-0.09**	**-0.15**	**-0.48**	Mean	705

Standard Deviation		Methods				
J&J	C&P	Local	MOS	SSFM	Station	Obs
1.76	1.60	1.24	1.67	1.44	**Belfast**	156
1.55	1.61	1.37	1.70	N/A	**Leeds**	194
2.51	2.64	1.74	1.84	1.76	**Lossiemouth**	67
1.48	1.63	1.30	1.93	1.47	**Birmingham**	288
1.83	**1.87**	**1.41**	**1.79**	**1.56**	Mean	705

Table 1 Errors for maximum temperature forecasts

An important measure of performance is the number of gross errors. While the aim of the forecast is to be as close as possible to reality, in practise the forecaster aims to get within +/- 2 °C of his goal. Using all stations combined, table 2 shows percentages of occasions when the methods and the model products were +/- 3 °C or more in error. Note that these occasions are independent of each other.

	Obs	Occasions	%age
J&J	614	147	23.9
C&P	616	207	33.6
Local	625	35	5.6
MOS	640	78	12.2
SSFM	386	26	6.7

Table 2 Percentage occurrence of poor forecasts by method for all stations

Taking these results at face value it can be said that J&J is a better method than C&P if the forecaster's aim is to minimise his errors.

Table 2 should be good news to all practising forecasters. Assuming a 'bad' forecast of 3°C or more out, the forecaster relying entirely on the performance of C&P would find he was wrong more than 1 in 3 times whilst MOS would bring him down to about 1 in 10. This is further improved through the 'local' method and the SSFM which only get it wrong about 1 in 16 attempts.

Whilst gathering the data, the forecasters were asked to enter an airmass/wind direction description. This took the form of a Lamb Classification – a single letter to define the airmass of the day i.e. F was 'cyclonically curved E'ly'. At the start of the project this extra data was considered to be worth having and possibly producing an answer along the lines of, 'use Bloggs method in a SW'ly but Baggins in a N'ly'. With 28 different classifications being possible insufficient data precluded such an approach but a simpler method was tried by dividing up the data into three groups.

Group 1 – Anything from 180° to 270° - This being roughly mild, moist maritime air and mostly stable
Group 2 – Winds between 270° to 360° - Cool maritime and mainly tending towards unstable
Group 3 – Anything else from an E'ly quarter – Could be stable/unstable, moist or dry but having a continental origin.

Obviously, these are very rough and ready classifications and open to argument but an analysis was performed on the data for each station. The results for three stations (*Appendix 3*) confirmed the belief of forecasters that some situations are easier than others to get right. For example the RMS error for Group 2 directions at Belfast was almost twice as high as for Group 3 situations. However, this pattern followed through to all the methods as well as MOS/SSFM. There didn't appear to be any convincing evidence to show that method A is more advantageous over automated products than B or visa versa in a particular weather situation, thus, this line of investigation was not pursued any further.

The final approach was to look at scatter diagrams and correlations. An example scatter diagram for one station (*Appendix 4)* shows the correlation for McKenzie against the actual minimum temperature. Comparing correlation coefficients between stations was considered. Obviously, this would show a fairly high coefficient but to draw any conclusion from the actual values between stations was not considered to be of any great usefulness and this line of investigation was not pursued further.

3.2 Minimum Temperature

The 4 stations that took part in this trial were, Glasgow, Coningsby, Wittering and Benson. Three well known and popular traditional methods were assessed, McKenzie, Craddock and Barthrum. (*Full details in FRB at 2.4.2.1, 2.4.2.2 and 2.4.2.3 respectively)*. The methods are all very different in their approach but require forecaster input to judge overnight cloud amount and windspeed. This in itself could make the argument that the method maybe is very good but the action of the forecaster inserting misjudged parameters devalues it. However, since all the methods require these parameters at the same time, even if they are wrong, it is fair to judge method against method.

Looking firstly at table 3

RMS Error			**Methods**				
McKenzie	Craddock	Barthrum	Local	MOS	SSFM	Station	Obs
1.87	2.09	1.57	N/A	1.64	1.80	**Glasgow**	199
1.95	2.22	1.92	N/A	1.48	1.95	**Coningsby**	135
2.70	2.82	2.23	2.43	2.31	2.35	**Benson**	117
1.84	2.22	N/A	1.24	1.52	1.72	**Wittering**	184
2.09	**2.34**	**1.91**	**1.84**	**1.74**	**1.95**	**Mean**	635
Mean Error			**Methods**				Obs
-0.40	-0.69	0.11	N/A	-0.11	0.31	**Glasgow**	199
-0.03	-0.05	-0.47	N/A	-0.11	-0.03	**Coningsby**	135
-0.09	-0.29	-0.31	-0.36	-0.09	0.12	**Benson**	117
-0.70	-0.98	N/A	-0.26	-0.31	-0.81	**Wittering**	184
-0.30	**-0.40**	**-0.22**	**-0.31**	**-0.15**	**-0.10**	**Mean**	635
Standard Deviation			**Methods**				Obs
1.83	1.98	1.58	N/A	1.64	1.77	**Glasgow**	199
1.95	2.23	1.87	N/A	1.48	1.95	**Coningsby**	135
2.70	2.82	2.21	2.42	2.32	2.36	**Benson**	117
1.70	2.00	N/A	1.22	1.49	1.53	**Wittering**	184
2.05	**2.26**	**1.88**	**1.82**	**1.73**	**1.90**	**Mean**	635

Table 3 Errors for minimum temperature forecasts

Three stations are very similar in their results but Benson stands out as showing nearly a degree error worse than the others. The magnitude of this error is roughly similar for all of the traditional methods and is also seen in the model products. The fact is, Benson, as most forecasters know, is a tricky place to forecast night mins!

Out of the three methods Craddock was consistently the worse performer whilst Barthrum appeared best. Rather disturbingly the 'local' method at Benson was slightly worse than Barthrum. This possibly suggests that forecasters there look at McKenzie and MOS and apply their own adjustments when in fact results show that Barthrum without any adjustment is the best performer. Overall, however, the model products MOS and SSFM appear to perform better than the traditional methods.

Table 4 shows the occurrences of gross errors at Benson, that is, errors of ≥±3 °C, and table 5 shows the same for all stations.

	McK	Cradd	Barth	Local	MOS	SSFM
Obs	111	111	109	111	111	105
3°C err	27	30	18	20	19	22
%age	24.32	27.03	16.51	18.02	17.12	20.95

Table 4. Percentage occurrence of poor forecasts at Benson

Method	Mck	Cradd	Barth	Local	MOS	SSFM
Obs	551	552	350	386	546	359
3 °C err	85	104	40	37	68	58
%age	15.43	18.84	11.43	9.59	12.45	16.16

Table 5. Percentage occurrence of poor forecasts by method for all stations

As with the results for maximum temperature, it appears that using all stations the lowest number of poor forecasts arise using the 'local' method. Again this will provide comfort to forecasters as it demonstrates that they are able to add value to the forecast.

Table 6 shows the results for all stations combined using the two traditional methods and MOS for the same number of observations. The smallest percentage ≥±3 error is clearly shown to be MOS

Method	McK	Cradd	MOS
Obs	541	541	541
3 °C err	84	100	68
%age	15.53	18.48	12.57

Table 6 Percentage occurrence of poor forecasts for the 3 most commonly used techniques for all stations when data were available from all three methods

3.4 Fog formation and clearance

Having discussed max and min temperatures, conclusions and ideas suggested from the remaining trials all suffer from a lack of data due to the fact that we are dealing with relatively rare events. However, using what data was gathered, some conclusions were derived but they must be treated with extreme caution. Fog forecasting is notoriously difficult. The fog point for the airmass may be perfectly predicted but other factors can conspire to make a

nonsense of it. For instance nearby moisture sources, rivers etc, can result in a much higher fog point than the method predicts. Also fog does tend to 'move around' with expected clearances delayed or at times coming much earlier.

Due to the many, (often unmeasurable), variables involved I confined the trial to radiation fog only. 4 stations took part for both formation and clearance of radiation fog. The trial started in November – just as an almost unbroken period of strong winds and depressions continued to affect UK throughout the winter! Notwithstanding this, a total of 27 fog formation events were gathered from the stations involved and an analysis of the data was made, however out of these events some methods and products were not available or recorded.

3.4.1 Fog Formation

The two traditional methods that appear in the FRB are Saunders technique and Craddock and Prichard's Method. Both methods predict a fog point temperature – the former using a representative T-phi and the latter based on a regression formula. *(Details in FRB 3.3.3.1 and 3.3.3.2)* The test of their performance was simply, given a radiation fog situation, compare the predicted fog point to the actual temperature when fog occurred. Forecasters were also asked to look at the SSFM and determine the temperature (fog point) when the model predicted fog formation. This produced the following result in Table 7a

	Saunders	Craddock	SSFM		Saunders	Craddock	SSFM
Obs	25	26	14		13	13	13
RMS	1.81	2.34	1.14		2.02	2.46	1.13
M Error	-0.36	0.76	-0.04		-0.75	0.42	0.05

Table 7a Fog point errors for the Saunders and Craddock methods and for the SSFM. LHS shows all obs and RHS of table shows methods and SSFM for the same cases.

The above table gives strong evidence that the SSFM produces a superior fog point prediction than either of the manual methods and Saunders shows a slight superiority over Craddock.

Using the same observations the distribution of the predicted fog point errors for Saunders and Craddock against the observed fog point is shown on Figure 1. Given such small samples very little can be deduced with any degree of confidence, however, there does appear to be a bias with Saunders appearing to produce a lower predicted fog point than was observed with Craddock showing the hint of an opposite bias.

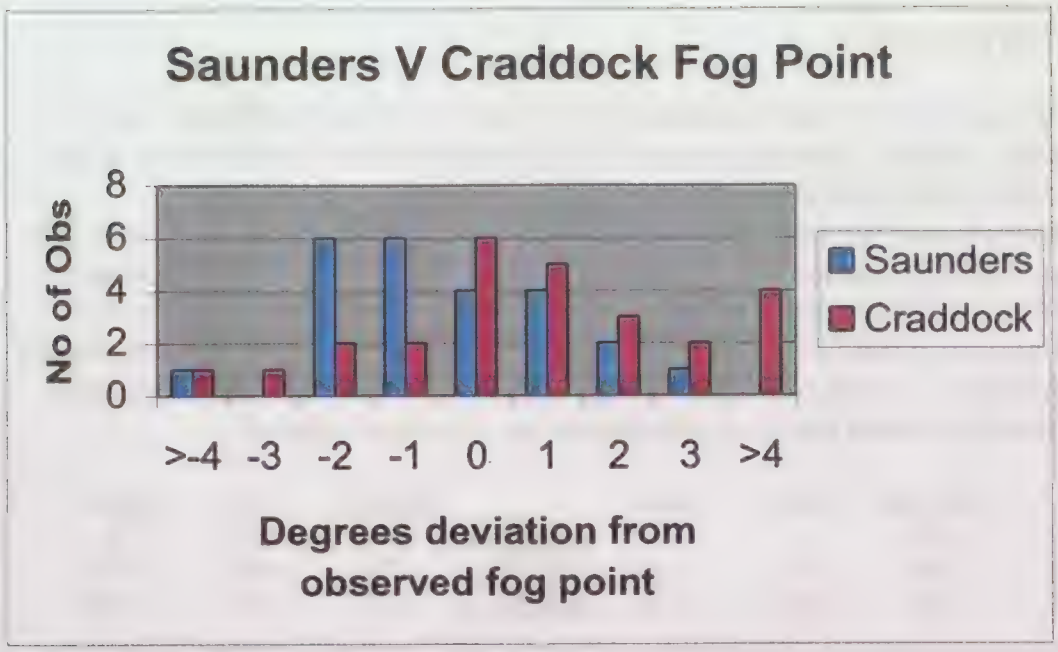


Figure 1

The range of error for the fog points were ±3°C for Saunders and ±4°C for Craddock. However, the SSFM with only 14 events produced a range of fog point error of +3 to –1 °C . A small sample, but results suggest that Saunders is a superior method for forecasting a fog point than Craddock, albeit with evidence to suggest Saunders shows some negative bias in it's fog point prediction. From the survey Saunders was by far the most popular a method and is much easier to use than Craddock. However, Saunders does depend of the availability of a suitable ascent. The SSFM produced a superior result than both traditional methods, but with a smaller sample size. This gives fairly convincing evidence that the lack of a suitable ascent does not prevent the forecaster from determining a fog point. It certainly indicates that given the choice between Craddock and SSFM he would be advised to choose the latter.

3.4.2 Fog Clearance

For fog clearance the two traditional methods tested were Jefferson and Barthrum. Strictly speaking these are not independent of each other – both use a representative ascent but the former predicts a fog clearance temperature and the later the time of clearance. *(Details in FRB 3.3.4.1c/d)* The method used to analyse the 16 recorded clearance events was to compare the clearance temperature and time with the two methods along with the prediction from the SSFM as well as interpretation from the Mesoscale model. Looking, firstly, at the predicted temperature clearance from Jefferson and from the two model products the results are shown in table 8

	Jefferson	Meso	SSFM		Jefferson	Meso	SSFM
Obs	16	9	8		6	6	6
RMS	1.90	1.12	1.06		2.79	0.96	0.97
ME	0.89	0.37	-0.13		1.13	0.55	-0.45

Table 8 Fog clearance temperature errors for Jefferson and model products. LHS shows all obs and RHS of table shows methods and SSFM for the same cases.

Again, very small sample sizes, but the results indicate that, for the forecaster, the SSFM or the Meso seem the best starting points to answer the question of clearance temperature. Both model products produced an error range of –1 to +2 °C whilst the corresponding Jefferson range was –1 to +4 °C. Simply by knowing the likely temperature when the fog should clear is not a great deal of help, however, since the Customer is more likely to want to know when it will clear!

Table 9 shows the results for fog clearance time errors – these expressed in hours.

	Barthrum	Meso	SSFM		Barthrum	Meso	SSFM
Obs	9	10	8		5	5	5
RMS	3.73	2.66	4.39		2.57	1.43	5.01
ME	2.82	2.28	3.21		1.70	-0.99	3.45

Table 9 Fog clearance time errors(hrs) for Barthrum and model products. LHS shows all obs and RHS of table shows methods and SSFM for the same cases.

In practical terms of providing good guidance to a Customer these figures are not very encouraging. The worst errors showed Barthrum 7 hours out and SSFM on one occasion 10 hours out. The Meso fared a little better with the worst performance being 4 hours out, this significantly, occurred when SSFM was 10 hours adrift. It has to be stressed, however, that with such a small sample size no definite conclusions can be made based on these results. There is no doubt that the basic physics and logic behind the methods are sound but local effects and, for example, medium and upper cloud arriving over the top of the fog layer can totally change the clearance (or perhaps non-clearance) time. In fairness to the methods, to be able to accurately forecast clearance time is even more difficult than fog formation due to the numerous factors involved.

Although a very small amount of data was obtained there looks to be some evidence to claim that the methods and, perhaps to some extent, the model, are not a great deal of help to the forecaster in terms of timing. In a run of radiation fog nights the forecaster can be fairly confident of getting the formation and clearance times right if there has been no change in the synoptic situation and can also make adjustments for perhaps somewhat drier/moister air or a little more or less wind. However, in most situations he is on his own and must make judgements based on experience. Satellite imagery seems to offer some hope for the future by use of images at night from the Polar Orbiter in the Channel 3/4 wavelengths where fog is emphasised. During daylight, visible imagery is better. At all times, this information is incorporated in the Nimrod visibility analyses and forecast, which should therefore be the preferred source of actuals and nowcast guidance. If possible looping or comparing pictures can enable the forecaster to see where the fog is and get a feel for the speed of clearance.

3.5 Thunderstorms and Associated Gusts

This part of the trial was probably the most difficult to obtain useful data, due to the localised nature of thunderstorms and associated gusts. No doubt many events were missed but with 33 examples obtained the data has some degree of significance. The method used was to test the various thunderstorm forecasting techniques, these being instability thresholds, and also to combine this with the expected maximum gusts using the Fawbush and Miller Technique and that predicted from the SSFM. Given an expected thundery situation, forecasters were asked to inspect the most representative T-Phi for their station and enter on their spreadsheet the numbers the various instability indices produced.

3.5.1 Results testing Instability Indices

In this trial forecasters were asked, given an expected thundery occasion, to record the instability indices from the traditional methods. (*Details of the methods are described in FRB 4.7.2*). 33 thunder situations from a total of 4 stations were analysed. There may or may not have been a thunderstorm at each station in every case but in all cases comments indicated that either SFLOC's were occurring or there were reports of storms nearby. Results are shown in table 10

Methods		Hit Rate
Boyden	All >=95	100
Rackliffe	All>=29	100
Jefferson	7 below limits	78
Pot Inst – Winter	Correct	0
Pot Inst – Summer	2 Right	12
K Index	23 Right	72

Table 10 Hit rate for thunderstorm (on-site or nearby) prediction by method for all stations

Boyden and Rackliffe proved correct in all events. However, as the analysis only looked at actual thunder situations, it was not possible to determine how many false-alarms these methods would have given. Jefferson, which is an amended and supposed improved version of Rackliffe, was correct in only 78% of events. The Potential Instability Index appeared to be a totally inadequate prediction technique, whilst the K index proved correct in 72% of occasions.

Although a relatively small sample the trial appears to show that Boyden and Rackliffe are the best predictors of thundery activity and that forecasters need pay no attention to the other methods. The forecast problem of accurately predicting the time and location of a thunderstorm more than a few hours ahead still remains intractable. The experienced duty forecaster will be aware of the potential for thunderstorms by knowing the airmass and inspecting the relevant T-Phi. With the help of SFLOC reports, satellite loops and radar he can now give good guidance to customers up to an hour ahead.

Whilst instability indices have been useful in the past, with the data available nowadays, they are of limited use to the practising forecaster and there appears to be no need to introduce a new forecaster to them.

3.5.2 Results testing Gusts associated with Thunderstorms

The chances of an extreme convective gust front event being recorded by a station anemo must be very small but we know by experience that they do happen and it would be of great use to the forecaster to be able to at least warn his customers of the possibility and magnitude of a damaging gust in the vicinity. There will always be a 'crying wolf' tendency due to the nature of the problem but within a thunderstorm warning a mention of the possibility of an extreme gust is always helpful.

Fawbush & Miller describe a technique that relates peak gust strengths in a thunderstorm based on the negative buoyancy force in a downdraught. The gusts are related to the temperature difference between the drowndraught air and the surrounding air near the surface. In the trial forecasters entered

the actual fall in temperature during the thunder event along with the derived Fawbush & Miller temperature fall from the relevant T-Phi. An approach to test if a relationship between the two exists is to calculate the correlation coefficient. The correlation coefficient ranges from 0, no relationship whatsoever between the two variables to 1, a perfect relationship. There exist various statistical techniques to quantify the coefficient but in this trial a simple comparison of them was considered appropriate. In this case a simple plot, figure 2, of the downdraught depression against the max gust recorded using 26 events gives a correlation of 0.56 and a casual glance at the graph seems to show a weak relationship.

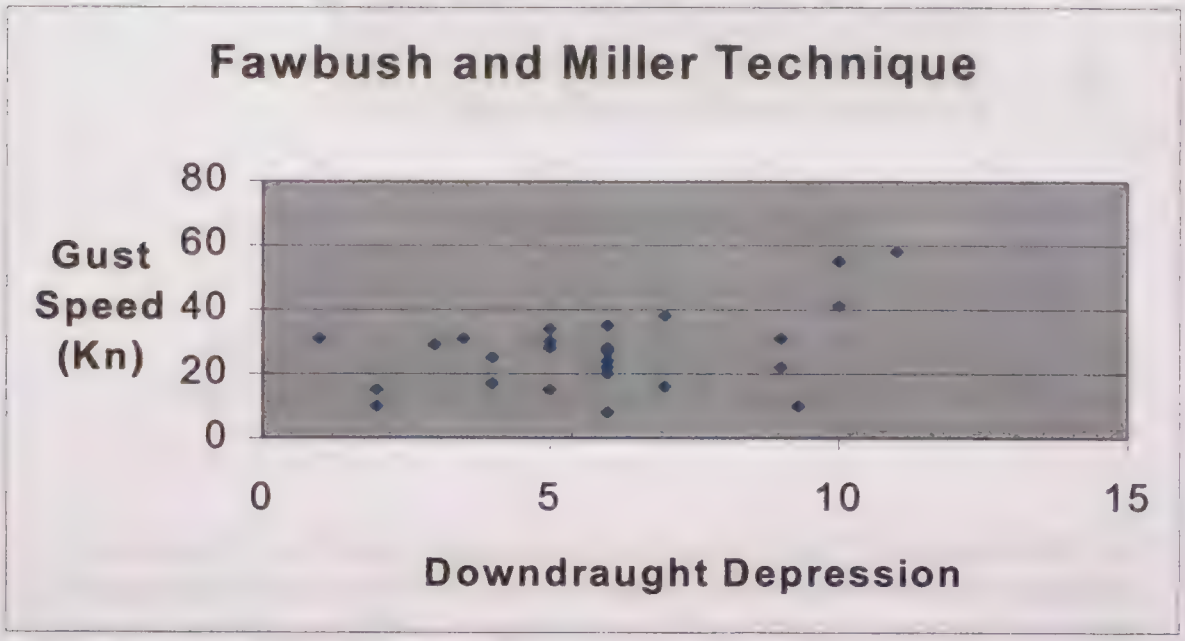


Figure 2. Showing the relationship between gust speed (kn) and thunderstorm downdraught depression (C).

However, to be fair to Fawbush and Miller, the method describes the maximum gust likely to be encountered on the surface below the thundercloud. This often is not recorded at the station but can occur perhaps half a mile away.

Assuming the station did capture an extreme gust the relationship looks rather more convincing. By excluding those events when no gust of over 25kt occurred it is reasonable to accept that those remaining did capture a gust associated with the thundercloud. The results shown in figure 3 for 14 gust events now shows a stronger correlation of 0.74

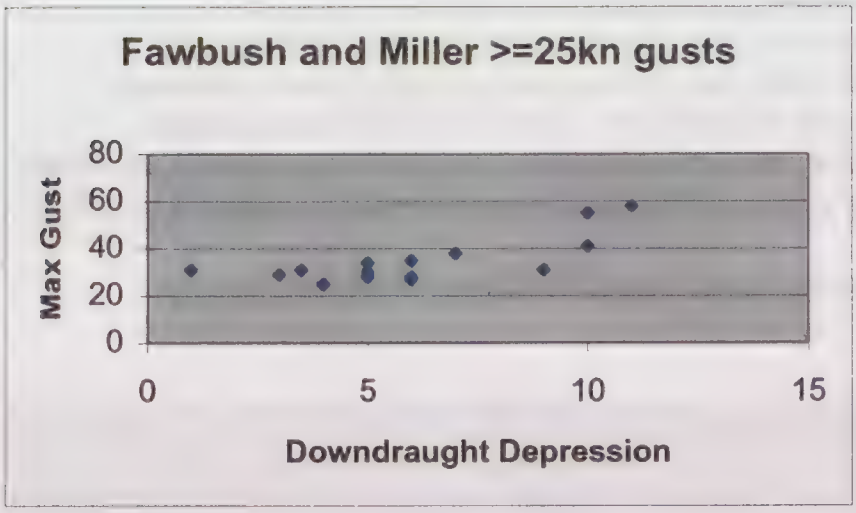


Figure 3. Showing the relationship between gust speed (kn) and thunderstorm downdraught depression (C) for events where a gust of >=25kn was recorded.

The SSFM output includes a 'max convective gust' predictor. 12 events were available for analysis. The SSFM maximum gust predictions did not inspire great confidence. On one occasion a max gust of 140kn was predicted when the station reported a 'gust' of 8kn! However, using the same reasoning as before, a plot of SSFM forecast gusts against actual gusts of >=25kn was produced (Figure 4). This gave a correlation coefficient of 0.54.

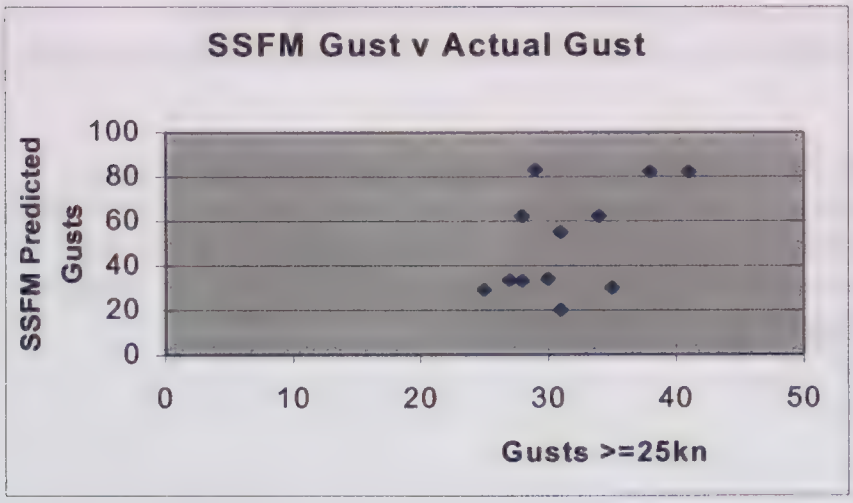


Figure 4. Showing the relationship between gust speed (kn) and thunderstorm downdraught depression (C) for events where a gust of >=25kn was recorded.

Given the erratic quality of the convective gust output from SSFM, forecasters should certainly not accept the figure as gospel. Perhaps the best course of action when a large gust is predicted is for the forecaster to take a look at the Fawbush and Miller technique to check the magnitude of the downdraught depression. If this turns out to be more than about 5 degrees then it would seem prudent to at least mention the possibility of gusty winds in the vicinity of a thunderstorm.

3.6 Snow

The forecasting of snow in the UK is always fraught with problems. Whilst an unexpected outbreak of light rain is not usually too serious to customers the moment a snowflake is noticed there is general excitement and the potential for embarrassment to the forecaster. Generally speaking snow showers are usually forecast quite well, with the main problem being their distribution and intensity, however, the 'will it, won't it', snow situation when frontal systems approach from the SW during a cold spell cause the forecaster the most problems. In the snow trial I confined the study to the latter situation.

The FRB mentions 7 methods to predict snow (*FRB 5.10.1.1 to 5.10.1.7*). These methods are generally developed on the obvious factor of determining if the air is cold enough to produce snow reaching the surface and are expressed in terms of percentage probability or in purely subjective expectations. Based on a mixture of their popularity and the degree of their different approaches 4 methods were tested along with model products.

Boyden looks at the 1000-850hPa thickness whilst another method, the height of the 0°C wet-bulb, predicts four likely outcomes of snow or no-snow. Booth and Varley's methods simply consider the temperature and dew point at the surface to predict the likelihood of snow.

Although news reports during the 2000/2001 Winter seemed to show frequent occasions of snow ploughs and abandoned vehicles, the 6 stations chosen for the trial seem to have missed many falls and in total only 18 potential snow events were recorded. It may well be that one snow predictor is better than another at a particular station, but given the paucity of data no attempt was made to investigate this. Thus, the data from all of the stations was combined and limited conclusions were made. Given an expected snow event forecasters were asked to enter the %age probability of snow falling using Boyden, the height of the web bulb freezing level, Booth's and Varley's methods followed by model output from the SSFM and the PGM. After the event a yes/no answer was entered if the snow occurred and also an indication of the intensity of the snow.

No attempt was made to derive any statistics from the data but the following points were made (and open to discussion). The first difficulty to resolve was the fact that the methods predict the likelihood of snow reaching the ground

if precipitation were to occur. Thus, the data was restricted to only the cases where snow actually did occur or where the models were predicting snowfall.

Looking firstly at Boyden, results were very variable. Without attempting any statistical analysis the results seem to indicate that a Boyden figure of over 75% is a good predictor generally, however 4 snow events out of the 18 recorded gave Boyden figures of 20% or less.

The wet-bulb height, not surprisingly, shows, without exception, that precipitation fell as snow when it was lower than 1000ft. Similarly, above 3000ft no snow events were recorded. Neither, however, is any great help to the forecaster in a marginal situation.

Booth's method is a straightforward probability of snow reaching the ground based on half the sum of the dry bulb and dew point. Thus, the formula produces three categories – mostly snow, snow and rain mixed, and mostly rains. The results seem to bear this out reasonably but again the method doesn't provide the panacea to the marginal problem.

Varley was only used at Cardiff and little can be concluded from the results.

Table 11 shows the results for the 18 snow cases

Wet Bulb height:- Enter 4 for >3000ft 3 for 2000-3000ft 2 for 1000-2000ft and 1 for <1000ft								
Cardiff					Models			
Methods	Boyden	Wet Bulb	Booth	Varley	SSFM	PGM	Actual	Intensity
Date	%	4,3,2,1	1/2(T+Td)	Melting ht	y/n	y/n	y/n	l/m/h
15-Dec-00	10%	3	2	100m	N	y	y	l
23-Dec-00	15%	4	3.5	200m	n/a	y	n	
24-Dec-00	90%	3	3.5	200m	n/a	y	y	l
27-Dec-00	90%	1	-2	surface	n/a	y	y	h
30-Dec-00	75%	1	0.5		n/a	y	y	m
20-Jan-01	75%	1	0		n/a	n	y	m
25-Feb-01	35	2	2	surface	N	n	y	m
16-Mar-01	30	2	0.75	surface	n/a	n	y	m
Linton					Models			
	Boyden	Wet Bulb	Booth	Local	SSFM	PGM	Actual	Intensity
Date	%	4,3,2,1	1/2(T+Td)	y/n or %	y/n	y/n	y/n	l/m/h
01-Feb-01	20	1	0.5		N	n	y	l
25-Feb-01	50	1	-0.4		N	n	y	l
28-Feb-01	80	1	0	y	N	y	y	m
Wattisham					Models			
	Boyden	Wet Bulb	Booth	Local	SSFM	PGM	Actual	Intensity
	%	4,3,2,1	1/2(T+Td)	y/n or %	y/n	y/n	y/n	l/m/h
19-Jan-01	70	1	-0.5	100%	N	n	y	l
26-Jan-01	25	3	2.5	Rain/snow	Y	n	y	m
28-Feb-01 am	45	1	1.7	Rain/snow	N	y	y	m
28-Feb-01 pm	60	1	1	Rain/snow	nil sig	y	y	
20-Mar-01 pm	50	2	1	Rain/snow	Y	y	y	
21-Mar-01 am	25	1	1	snow	N	y	y	m

Table 11. Forecast probabilities using the different methods for the 18 snow cases.

Looking now at the model forecasts. The PGM provides a broader picture of the distribution of snowfall and, in the trial, forecasters were asked to simply interpret whether a snow symbol appeared over their station. This is a very subjective test and it would be unfair to judge the PGM performance on this. The SSFM, however, by definition is a forecast for the specific site and the results show it to be less than distinguished. Out of 9 snow events the SSFM only predicted 2 cases of snow.

The data in this trial are plainly inadequate to come to any clear conclusions. In particular, independently derived thresholds for the traditional methods

were not available. Even with thresholds optimised for the trial period, they are not particularly useful to the forecaster in marginal situations.
In a communication with a member of the Forecasting Development team a view was expressed that SSFM does underpredict snow but there is insufficient evidence to support it.

3.7 Low Cloud

The FRB contains no clear methods for forecasting the height of low cloud. A series of bland statements concerning the physical conditions when low cloud will occur is of no use to a forecaster who is posed the question by a customer, 'So, how low will the Stratus be tonight?'. For most forecasters at a particular station the answer is based on their experience. On the whole, in the common warm sector situation, consideration of the airmass dew point is the prime factor along with ground temperatures in Winter. There is also an intuitive thought that windspeed is a factor involved in the height of low cloud – this based on an idea that perhaps the stronger the wind the higher the cloud is lifted from the surface. It was decided that this would be tested in the trials. Forecasters were asked to enter occasions, once a day at a fixed time, when stratus was occurring at their station with a nominal height of 1200ft or less and amount of 3/8ths or more. The windspeed, temperature and dewpoint were recorded and also the figure for cloudbase taken from the SSFM at the time was entered. Data was received from 4 stations, Valley, Kinloss, St Mawgan and Manchester. Again, due to a surprisingly small amount of data they were combined for analysis.

Two scatter diagrams were produced. The first in table 12 shows the actual cloud base against the windspeed and the second in table 11 shows how the model forecast compared with the actual cloudbase. The number of obs used were 32 for cloudbase v windspeed and 30 for cloudbase v model cloudbase.

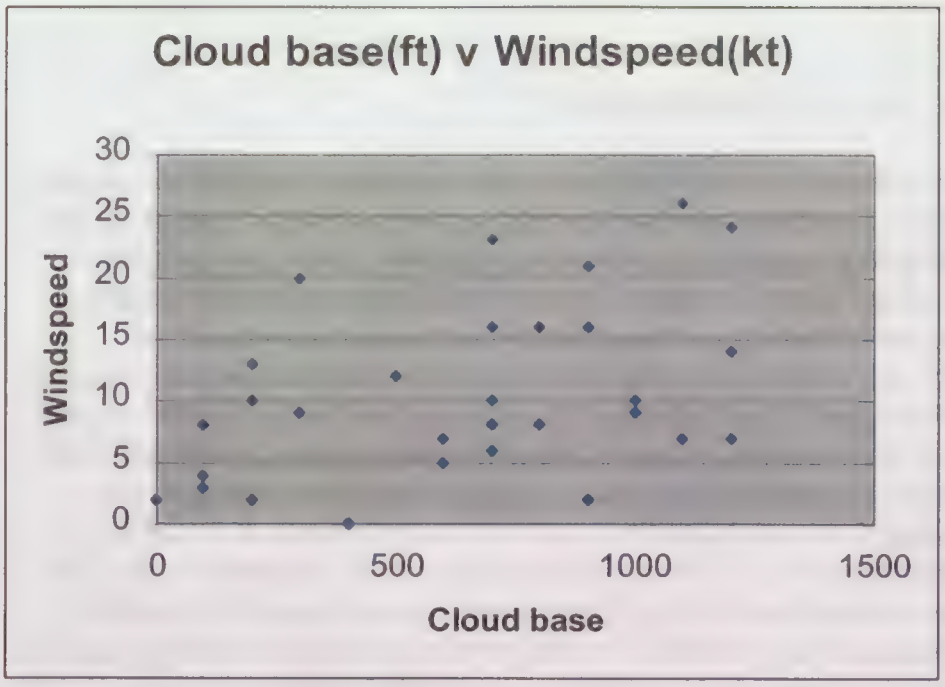


Figure 5. Relationship between cloud base and surface wind speed.

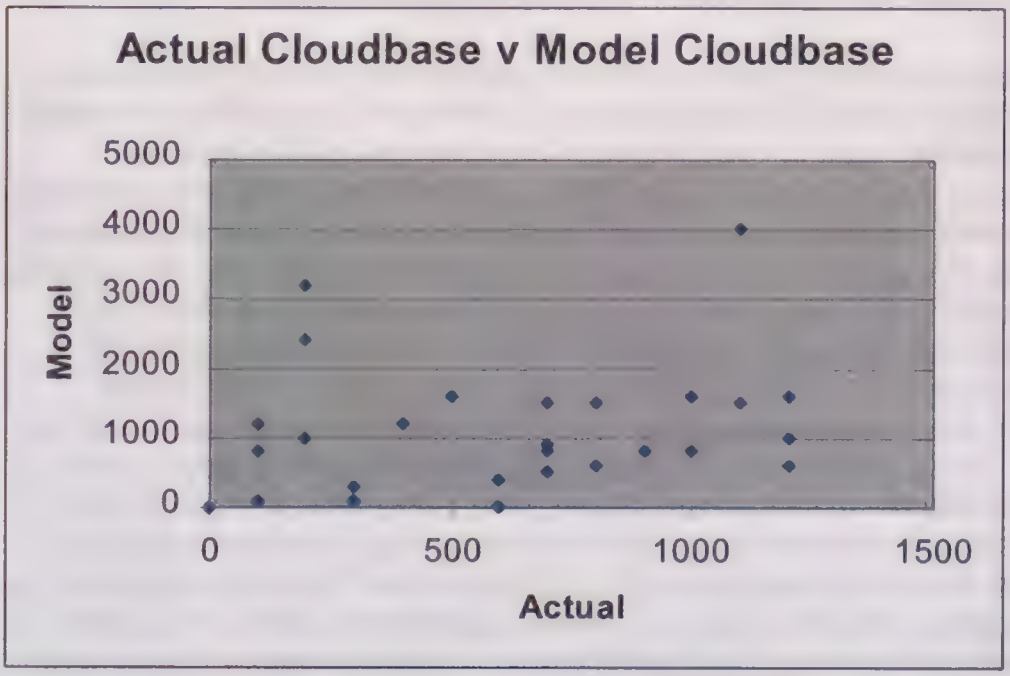


Figure 6. SSFM forecast and actual cloud base.

The two diagrams show at a glance that there appear to be no clear relationships. The first had a correlation of 0.46 whilst the second only managed 0.19. Neither provides any great assistance to the forecaster. The performance of the SSFM cloudbase certainlyappeared at first sight to be not very impressive, however, practising forecasters must be aware that the

model cloud output is in a probabilistic form. and needs to be carefully interpreted.

4. Conclusions and Recommendations

Having been a forecaster for twenty years I am well aware of those occasions when the forecast is perfect and the weather somehow doesn't quite behave itself! It is galling to give one's best efforts at getting it right only to have everything fall apart just at handover after a twelve hour shift. Any help the forecaster gets, be it from the old traditional methods or the model, is welcome. Nobody can deny that the model has been steadily improving over the years and we accept that the emphasis is firmly on the human forecaster adding value to the machine product. Many of the methods described in the FRB are nowadays superfluous but they do give an insight as to how pre-computer product forecasters performed their task. In terms of the aims of the Project some quite clear conclusions can be made, some debatable and for the weather elements with only limited events really very little breaking ground.The initial survey to establish which methods were still being used by forecasters revealed that out of the large number of methods for night minima described in the FRB only two were used regularly – these being McKenzie for night mins and Saunders for fog points. The others were used in varying degrees depending on the frequency of the event.

Taking max and min temperatures firstly. By their very nature a substantial amount of data was able to be gathered – well in excess of 500 observations from the various stations. As previously described the results do show conclusively that the model output from both MOS and SSFM are every bit as good or even better than any of the traditional methods. There is also good evidence, though, that the forecaster does contribute some value by applying adjustments to the model values based on his experience and skill. The hypothesis that forecasters derive their 'local' method by using a mix of traditional methods and model products is, in itself , not unreasonable. Certainly, for an inexperienced forecaster the traditional methods should be used as a check on the performance of model products. It obviously gives more confidence to the forecaster if, say, McKenzie, MOS and SSFM are all coming up with the same figure for the forecast night minimum. However, on hopefully rare occasions, they may all be in error! The results do show that if a case arises when the manual method is significantly different from the predicted model forecast that the forecaster would do better in the long run to believe the model.

The formation and clearance of fog trial was disappointing in regard to the number of events obtained. However, there seems to be a reasonable indication that Saunders fared better than Craddock for a fog point temperature whilst results for clearance times from Jefferson and Barthrum were both very poor albeit based on a very small amount of data. For both trials, the SSFM gave better results than either of the traditional methods. The

only clear conclusion to be made is that the fog problem remains generally unresolved by either traditional methods or model products.

Thunderstorms and gusts were probably the most difficult trial to conduct for the lack of events and due to the relatively small area of effect. The traditional methods for producing instability indices gave good hit rates but they do belong to the past and today with good SFLOC coverage and satellite pictures they can be ignored. The Fawbush and Miller technique for forecasting associated gusts produced a reasonable correlation with strong (≥25 kn) gusts and used alongside the predicted max gust from SSFM the forecaster should be able to provide his customer with a least a good indication that damaging gusts are possible in his area though quantitative estimates are unlikely to be reliable.

The snow trial provided very little data and minimal insight. The traditional methods were reasonably good but do not provide any great help in a marginal situation.

Finally, the low cloud trial provided very little new information. Although there were weak correlations, no conclusions could be made given the paucity of data.

Forecasters continue to comment on the seemingly continuing decrease in the number of suitable ascents available to them. This study has demonstrated that methods which rely on an ascent are generally effective, however, the 'no ascent' methods are not totally ineffective and the performance of model products more than make up for the lack of observational data.

5. Implications for Training

Forecaster training has evolved over the years and much less emphasis is placed on the use of the traditional methods. The Project has shown that in some cases one method is better than another but has also demonstrated that forecasters could get along fine without knowing anything of their existence. However, it would be wrong to discard the methods entirely since inexperienced forecasters do need to have some alternative ideas and values to compare with the model output. Probably the best policy would be to introduce them to those that have shown to perform best in this project and leave the others for new forecasters to read about if they so choose.

6. Finale

This project initially seemed straightforward but as more thought was put into it more difficulties became apparent. Whilst not coming up with any earth shattering new concepts and discoveries I hope those who have read this paper will appreciate and understand that weather forecasting is not an easy business and as useful as the computer models are, we are still a long

way from dispensing with the need for the human forecaster. No matter how experienced a forecaster is, there is always a weather situation that can catch one out and I believe everybody does their utmost to minimise these unfortunate occasions. Communication amongst forecasters is essential and the idea that we are all on the same side must be fostered. The internet and newsgroups are an excellent way of passing on the latest weather information – i.e. 'rain has just turned to snow here at xxxx'. Newsgroups dedicated to this project are found in 'met-office.forecast.techniques' and there is an excellent external group at 'uk.sci.weather'.

Finally, I would like to thank Gordon Clark (Ex NMC forecaster) for his assistance and advice with this Project, the Project Board and scientific advisors and also the S Met O's and forecasters in the outfield who have contributed and supplied the data.

Appendices

1. Analysis of Review Project Survey

Number of replies:- 157
Number of Stations:- 40 out of possible approx 45

Popular Methods for the various weather categories

Temperature:- 80% regularly used McKenzie for calculating minimum temperature. Johnson & Jefferson's method for daytime heating was used regularly by 38% of respondee's.

Visibility:- 79% regularly used Saunders to calculate fog points. The temperature construction by Jefferson was used regularly by 34%

Cloud:- 62% used the factors to forecast Frontal Cloud as described in FRB 5.2.1 but this is not a method as such. No other cloud forecast methods were used significantly.

Snow:- 54% used the height of the zero degree wet bulb temperature with 37% using Boyden's method.

Convection:- The criteria described in FRB 4.7.3 for hail was used regularly by 32% and occasionally by 54% - perhaps highlighting that hail is not forecast very often. Boyden appears most popular at 22% for forecasting thunderstorms.

Wind:- For mountain waves Casswell was frequently used by 43% of respondents. Significantly, Shutts method which claims to be better was never used by 93% of respondents.

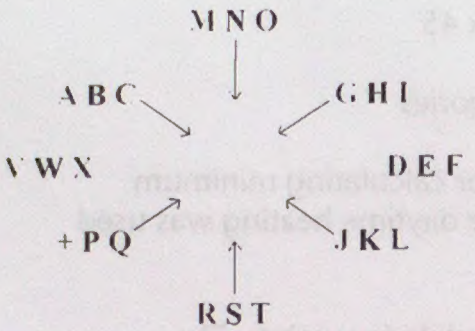
Initial conclusion:- Out of the numerous forecasting methods in FRB McKenzie's minimum temperature and Saunders fog point stand out as the only two that most forecasters regularly use.

Initial Interpretation of Survey:- A response rate of 85%+ must be considered as very good. Many of the methods involve constructions and nomograms. These may well have been used frequently in the past but in today's era the forecaster has very little time or perhaps inclination to use such methods. Whilst there may have been a 10% or so of respondee's who spent very little time in considering their marking of the survey I believe a large majority of forecasters gave it a fair degree of thought resulting in what I believe are valid results.

Appendix 2

Lamb Types

LAMB TYPES



Z cyclonic
U_c cyclonic col
Y anticyclonic
U_a anticyclonic col

USE OF LAMB DIAGRAM

The diagram is almost self explanatory in that there are three categories in each of the eight wind directions illustrated. The first in the group would be used for anticyclonic flow, the second for straight flow and the last for cyclonic flow.

Thus a cyclonically curved easterly flow has a Lamb classification of the letter 'F'. This classification may not be important for, nor relevant to, some of the techniques but nevertheless completion of this section was requested.

Appendix 3

Minimum Temperature RMS's & M Errors for Groups 1 to 3 wind directions

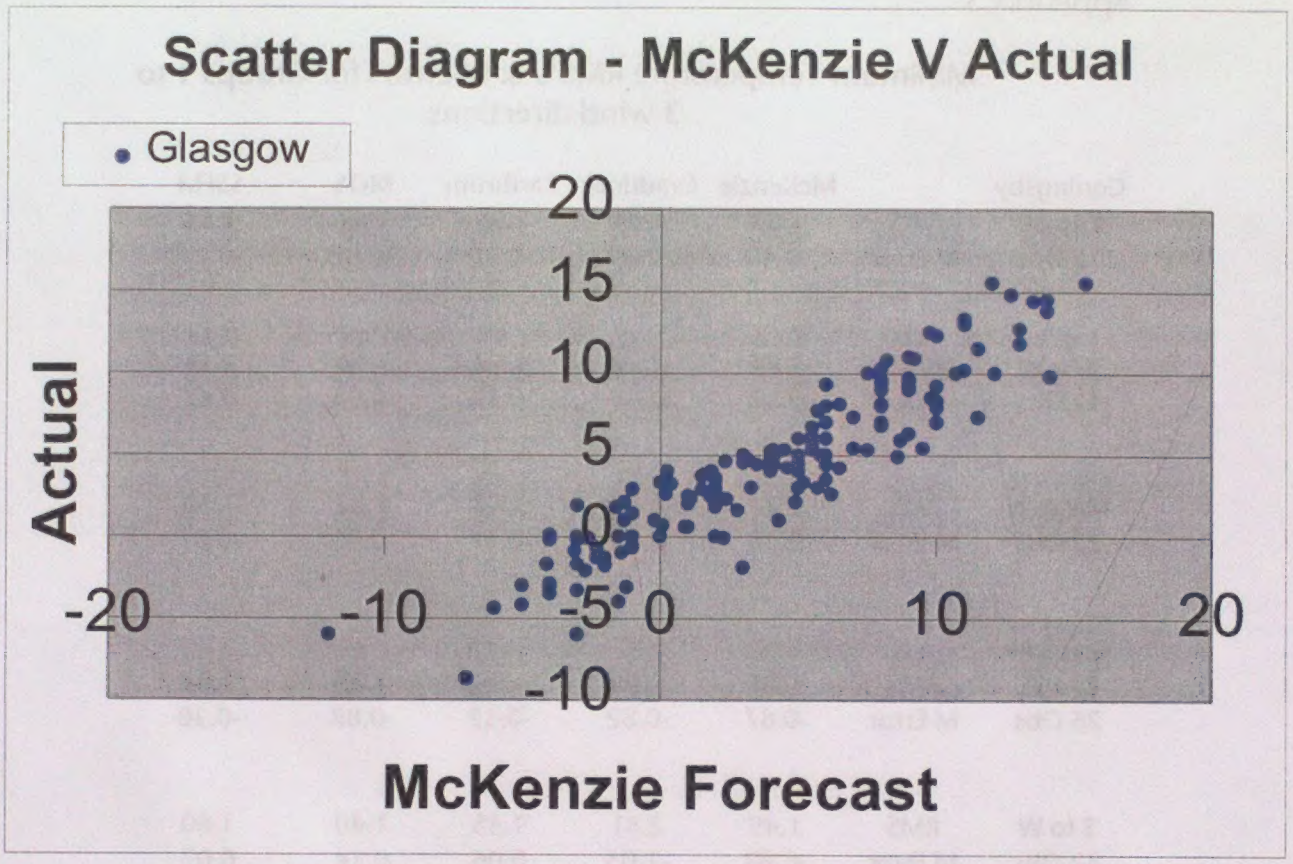
Coningsby		McKenzie	Craddock	Barthrum	MOS	SSFM
E to SE	RMS	1.83	1.94	1.96	1.51	1.54
15 Obs	M Error	-0.40	-0.04	-1.00	0.20	-0.17
S to W	RMS	1.94	2.22	1.97	1.64	2.16
41 Obs	M Error	0.50	0.64	-0.12	-0.05	0.42
NW to N	RMS	1.23	1.83	1.42	1.37	1.30
22 Obs	M Error	0.25	0.30	-0.37	-0.35	-0.07
Glasgow						
E to SE	RMS	1.71	1.95	1.94	1.87	1.84
26 Obs	M Error	-0.87	-0.82	-0.32	-0.87	-0.30
S to W	RMS	1.49	2.31	1.45	1.40	1.60
32 Obs	M Error	-0.48	-1.05	0.06	0.14	0.07
NW to N	RMS	2.58	2.01	1.77	2.08	1.71
14 Obs	M Error	-1.60	-1.32	-0.73	-1.52	-1.19

Maximum Temperature RMS's & M Errors for Groups 1 to 3 wind directions

Belfast

		J & J	C & P	Local	MOS	SSFM
E to SE	RMS	2.03	1.96	1.12	1.49	1.25
32 Obs	M Error	0.84	1.04	0.19	0.47	-0.51
S to W	RMS	1.91	1.40	1.07	1.25	1.34
35 Obs	M Error	1.14	0.78	0.15	-0.02	-0.19
NW to N	RMS	2.34	2.51	1.37	1.84	2.09
15 Obs	M Error	1.14	1.09	-0.16	0.44	0.28

Appendix 4



Minimum temperatures at Glasgow. Using 164 Observations.

Correlation Coefficient: 0.929